Journal of Education and Educational Research
ISSN: 2957-9465 | Vol. 1, No. 1, 2022

On the application of projection threading method in the teaching of probability theory

Kongsheng Zhang*

School of Statistics and Applied Mathematics, Anhui University of Finance and Economics, Bengbu 233030, China

Abstract: The probability density function of the sum of random variables and marginal density function are important and difficult concepts in the course of probability theory. The key to calculate them is to determine the range of interested variables and integral variables. Combined with graphics, the range of those variables can be determined quickly by using projection threading method. The proposed method is intuitive and easy to master.

Keywords: Projection; Threading method; Probability density; Sum of random variables; Marginal density.

1. Introduction

Random variable plays an important role in the course of probability theory and mathematical statistics. The additivity of random variables can be judged by the probability density function of the sum of random variables. Whether the random variables are independent can be judged by the relationship between the joint density function and the marginal density function. The famous convolution formula can be used to calculate the probability density function of the sum of random variables, but the formula needs to consider the range of two variables, one is associated to the sum, and the other is the integral variable. Although the range of these two variables is full plane, it may be necessary to properly compress this range for specific problems. The difficulty is how to determine the range of interested variables and integral variables. The range of two variables need to be considered similarly when calculating the marginal density. At the beginning of this course, students still have difficulty in the calculation process, and are still vague about how to solve these problems, although they have mastered the convolution formula and marginal density calculation formula.

Han and Xu [1] gave a simple method to determine the lower limit and upper limit of the integral, but the method is not easy to understand due to missing figures. When calculating the density function of the sum of binary random variables, Shi et al. [2] proposed an intuitive method which combined numbers with shapes to determine the upper and lower limits of integral variables, however, they only discussed the case of independence between random variables. Other papers also discussed the distribution of linear combinations of random variables [3-4]. Xie [5] used the projection threading method to convert double integrals into repeated integrals when calculating double integrals. Motivated by the above methods, we use the projection threading method more detailed to calculate the density function and marginal density function of the sum of random variables in the course of probability theory. This method mainly includes two steps: Firstly, the support set is projected to the axis corresponding to the variable of interest. Secondly, a straight line (horizontal or vertical) crosses the support set to determine the upper and lower bounds of the integral variable. In this paper, we plan to use the projection threading method to quickly determine the range of independent variables and the upper and lower limits of integral variables, simplify the integral calculation, and then obtain the probability density and marginal density expressions of the sum of random variables. This method makes full use of the information of graph, and the calculation process is simple and easy to master.

2. Calculation of Probability Density of Sum of Random Variables by Projection Threading Method

In this paper, we only consider the case of bivariate continuous random variables for brevity. Assume $f(x,y)$ is the joint probability density function of random variables (X,Y), $f_X(x)$ and $f_Y(y)$ are the maginal probability density functions for X and Y, respectively. Let $Z = X + Y$, by using convolution formula [6], the probability density function of Z is given as $f_Z(z) = \int_{-\infty}^{+\infty} f(x, z-x)dx$.

The above integral depends on the two variables x and z. Notice that the range of x and z is full plane, to ease the process of calculation, it is necessary to properly compress this range. Projection threading method can be used here to determine precisely the range of those two variables. This method facilitates the calculation for the corresponding probability density function. The projection threading method mainly includes the following four steps to calculate the probability density function of the random variable Z.

Step 1. Take the integral variable in the convolution formula as the horizontal axis and the independent variable as the vertical axis to establish a rectangular coordinate system (Note: the integral variable can also be used as the vertical axis, and the independent variable is the horizontal axis accordingly).

Step 2. Determine the support set $D = \text{support}(X,Z) = \{(x,z)|f(x,z-x) > 0\}$.

Step 3. Project the region D onto z-axis since the variable of interest is z, the projection area is recorded as A, and the outside area of A on the z-axis is represented by A^C.

Step 4. Threading. If $z \in A^C$, make a horizontal straight line through this point, it is easy to know that this line does not intersect with region D. Thus, for any $x \in R$, we have $f(x, z-x) = 0$, it follows that $f_Z(z) = \int_{-\infty}^{+\infty} f(x, z-x)dx = 0$. If $z \in A$, through the point z, plot a line perpendicular to the z-axis from left to right, obviously, the line has a common part with region D. For any $z \in A$, if the inlet line and outlet line do not change, the scope of z does not need to be subdivided. Otherwise, classified discussion is required. Dividing the region A into different disjoint regions A_1, A_2, …, A_m (Usually limited) in order to ensure that the inlet line and outlet line do not change in each region $A_i (i = 1,2,\ldots,m)$. And then determine the range of variable z,

The integral limit of the integral variable x is determined according to the entrance line and the exit line respectively. The expression of probability density function $f_Z(z)$ can be obtained by calculating integrals in turn.

Example 1 [7]. Assume the probability density function of two-dimensional random variable is

$$f(x,y)=\begin{cases}2-x-y, & 0<x<1, 0<y<1\\ 0, & \text{otherwise.}\end{cases}$$

Let $Z = X + Y$, calculate the probability density function $f_Z(z)$ for Z.

Answer: The density function is calculated according to the projection threading method.

We first establish a rectangular coordinate system with x being horizontal axis and z being longitudinal axis. Find support set $D = \{(x,z) | f(x, z - x) > 0\} = \{(x,z) | 0 < x < 1, x < z < x + 1\}$ which is illustrated in Fig.1. Project support set D onto the z axis, it is easy to know that the projection area is A=(0,2). Thus $A^C = (-\infty, 0] \cup [2, +\infty)$, for any $z \in A^C$ and $x \in R$, one has $f_Z(z) = \int_{-\infty}^{+\infty} f(x, z - x) dx = 0$.

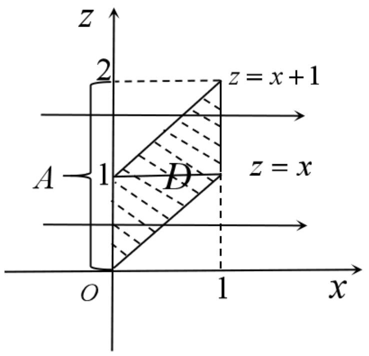

Fig.1 Illustration of projection threading method in example 1.

According to Fig.1, when $z \in (0,1)$, the horizontal straight line passing through the point crosses the area D with entrance line $z = 0$ and exit line $z = x$. However, if $z \in (1,2)$,

the entrance line becomes $z = x+1$ while the exit line becomes $x = 1$. Obviously, the exit line and the entrance line have changed. Therefore, the region A will be divided into two regions A_1 and A_2, where $A_1 = (0,1)$ and $A_2 = [1,2)$.

The case $z \in A_1$. According to the shape of the entrance line and the exit line, the density function can be calculated as follows

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{+\infty} f(x, z-x)dx \\ &= \int_0^z (2 - x - (z-x))dx \\ &= (2-z)z. \end{aligned}$$

The case $z \in A_2$. Similarly,

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{+\infty} f(x, z-x)dx \\ &= \int_{z-1}^1 (2 - x - (z-x))dx \\ &= (2-z)^2. \end{aligned}$$

Based on the above discussion, the probability density function of $Z = X + Y$ is

$$f_Z(z)=\begin{cases}(2-z)z, & 0<z<1\\ (2-z)^2, & 1\le z<2\\ 0, & \text{otherwise.}\end{cases}$$

Example 2 [3]: Assume random variable $X \sim U(0,1)$, $Y \sim Exp(1)$, X is independent of Y. Let $Z = X - 2Y$, calculate probability density function of Z.

Answer: Let $f(x,y)$ be the joint probability density function of (X,Y). The cumulative distribution function of $Z = X - 2Y$ can be calculated as follows

$$\begin{aligned} P(Z \le z) &= P(X - 2Y \le z) \\ &= \iint_{x-2y\le z} f(x,y)dxdy \\ &= \int_{-\infty}^{+\infty} dx \int_{\frac{x-z}{2}}^{+\infty} f(x,y)dy \\ &\xlongequal{y=\frac{x+t}{2}} \int_{-\infty}^{+\infty} dx \int_{-z}^{+\infty} f(x, \frac{x+t}{2}) \frac{1}{2} dt \\ &\xlongequal{t=-s} \frac{1}{2} \int_{-\infty}^{+\infty} dx \int_{-\infty}^{z} f(x, \frac{x-s}{2}) ds \\ &= \frac{1}{2} \int_{-\infty}^{z} [\int_{-\infty}^{+\infty} f(x, \frac{x-s}{2}) dx] ds. \end{aligned}$$

Thus, the probability density function of $Z = X - 2Y$ can be represented as

$$f_Z(z) = \frac{1}{2} \int_{-\infty}^{+\infty} f(x, \frac{x-z}{2}) dx.$$

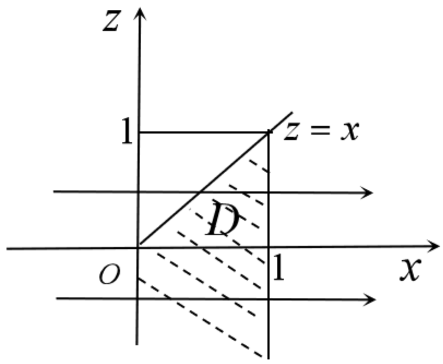

Fig.2 Illustration of projection threading method in Example 2

Then calculate the specific expression for $f_Z(z)$ by using projection threading method. Similar to Example 1, first establish a rectangular coordinate system, as shown in Fig.2. Then determine the support set $D = \{(x,z) | f_{X,Y}(x, z - x) > 0\} = \{(x,z) | 0 < x < 1, z < x\}$. And project the region D onto the vertical axis Z. The corresponding projection region $A = (-\infty, 1)$. Hence $A^C = [1, +\infty)$, for any $x \in R$ and $z \in A^C$, we have $f_Z(z) = \frac{1}{2} \int_{-\infty}^{+\infty} f(x, \frac{x-z}{2}) dx = 0$.

According to the characteristics of the projection threading method, it is easy to divide the projection area A into two parts, that is, $A = A_1 \cup A_2 = (0,1) \cup (-\infty, 0]$. According to Fig. 2, if $z \in A_1 = (0,1)$, any line passing through the point $(0, z)$ and perpendicular to the z-axis will go through the region D from left to right, the inlet line and outlet line are respectively are $x = z$ and $x = 1$, therefore,

$$\begin{aligned} f_Z(z) &= \frac{1}{2}\int_{-\infty}^{+\infty} f(x, \frac{x-z}{2})dx \\ &= \frac{1}{2}\int_{z}^{1} f_X(x) f_Y(\frac{x-z}{2})dx \\ &= \frac{1}{2}\int_{z}^{1} 1\cdot e^{-\frac{x-z}{2}} dx \\ &= 1 - e^{\frac{z-1}{2}}. \end{aligned}$$

If $z \in A_2 = (-\infty, 0]$, the corresponding entrance line is $x = 0$ and the outlet line is $x = 1$.

Hence,

$$\begin{aligned} f_Z(z) &= \frac{1}{2}\int_{-\infty}^{+\infty} f(x, \frac{x-z}{2})dx \\ &= \frac{1}{2}\int_{0}^{1} 1\cdot e^{-\frac{x-z}{2}} dx \\ &= e^{\frac{z}{2}} - e^{\frac{z-1}{2}}. \end{aligned}$$

Due to the above discussion, the probability density function of $Z = X - 2Y$ has the form

$$f_Z(z) = \begin{cases} 1 - e^{\frac{z-1}{2}}, & 0 < z < 1 \\ e^{\frac{z}{2}} - e^{\frac{z-1}{2}}, & -\infty < z \le 0 \\ 0, & \text{otherwise.} \end{cases}$$

3. Application of Projection Threading Method to Calculate Marginal Density

Assume the joint probability density function of (X, Y) is $f(x, y)$, then the marginal density function are $f_X(x) = \int_{-\infty}^{+\infty} f(x, y)dy$ and $f_Y(y) = \int_{-\infty}^{+\infty} f(x, y)dx$, respectively [6]. The calculation procedure of marginal density $f_X(x) = \int_{-\infty}^{+\infty} f(x, y)dy$ is similar to that of probability density of variable sum, however, the support set is $D = \text{support}(X, Y) = \{(x, y) | f(x, y) > 0\}$ here, the straight line crossing the area is perpendicular to x-axis. The calculation of $f_Y(y) = \int_{-\infty}^{+\infty} f(x, y)dx$ can be proceeded similarly.

Example 3 [8]: The joint probability density of binary continuous random variable (X, Y)

$$f(x, y) = \begin{cases} \frac{21}{4}x^2 y, & x^2 \le y \le 1 \\ 0, & \text{otherwise.} \end{cases}$$

Calculate the marginal density.

Answer: We firstly calculate the marginal density $f_X(x) = \int_{-\infty}^{+\infty} f(x, y)dy$ of random variable X.

(1) Different from Example 1 and Example 2, the horizontal axis of the coordinate system established here is variable x, the vertical axis still is variable y.

(2) In order to guarantee $f(x, y) > 0$, we only need the condition $-1 \le x \le 1$, $x^2 \le y \le 1$ and $(x, y) \ne (0,0)$ holds. Thus the corresponding support set is $D = \{(x, y) / -1 \le x \le 1, x^2 \le y \le 1\} \backslash \{(0,0)\}$ which is illustrated in Fig.3.

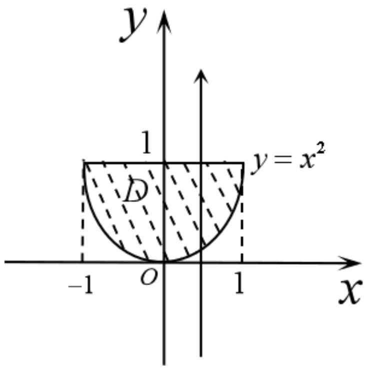

Fig.3 Illustration of projection threading method in Example 3

(3) We project the region D onto x-axis and obtain the project area A= [-1,1], it implies $A^C = (-\infty, -1) \cup (1, +\infty)$. For any $x \in A^C$ and $y \in R$, one obtains $f_X(x) = \int_{-\infty}^{+\infty} f(x, y)dy = 0$.

(4) When $x \in A$=[-1,1], take any point x in A, make a straight line parallel to the y-axis from bottom to top through the point x. This line will cross the region D with entrance line $y = x^2$ and exit line $y = 1$, Hence, $f_X(x) = \int_{x^2}^{1} \frac{21}{4}x^2 y dy = \frac{21}{8}x^2(1 - x^2)$.

Therefore, the marginal density function for random variable X is expressed as

$$f_X(x) = \begin{cases} \int_{x^2}^{1} \frac{21}{4}x^2 y dy = \frac{21}{8}x^2(1 - x^2), & x \in [-1, 1] \\ 0, & \text{otherwise.} \end{cases}$$

Similarly, the marginal density function for random variable Y has the form

$$f_y(y) = \begin{cases} \int_{-\sqrt{y}}^{\sqrt{y}} \frac{21}{4}x^2 y dx = \frac{7}{2}y^{\frac{5}{2}}, & y \in [0, 1] \\ 0, & \text{otherwise.} \end{cases}$$

4. Conclusion

The support set is determined according to the expression of the joint density function, and then project to the axis corresponding to the variable of interest. The range of integral variable is determined by threading method. The calculation of density function and marginal density function of variable sum are related to these two variables. The purpose of projection and threading is to determine the value range of these two variables respectively.

In this paper, the detailed calculation process of the density function of the combination of random variables is given intuitively by using the projection threading method, and the calculation of the marginal density function is also discussed. Through the analysis of three examples, it is shown that the method is easier to understand and master, and the range of the variable of interest and the integral variable can be determined accurately and quickly by combining with the graph. Then the corresponding probability density function can be obtained by using the ordinary integration technology. Projection threading method is also applicable to the calculation of density function of more general random variable combination.

Acknowledgements

The work is supported by Anhui Provincial Quality Engineering Teaching Research Project “Research on Teaching Reform of Probability Theory and Mathematical Statistics Based on Graphic Method"(No. 2020jyxm0015).

References

[1] Z.J.Han, Y.G.Xu: On the solution of the probability density function for the sum of two dimensional random variables, ShuXue XueXi yu YanJiu, (2014)No.21, p.77-77. (in Chinese).

[2] L.Shi,Q.Zhang, Y.M.Shi: On teaching method for the distribution of two dimensional random variable functions, Journal of Mudanjiang College of Education, (2018)No.7, p.42-44.(in Chinese).

[3] Q.Liu: Solving methods for the density function of sum of random variables, Journal of Science of Teachers’ College and University, Vol.37(2017)No.11, p.63-66.(in Chinese).

[4] W.B.Wang, X.B.Chang, M.Yu: Probability density of two dimensional continuous random variable linear combination, Studies in College Mathematics, Vol.18(2015)No.3, p. 23-24.(in Chinese).

[5] Q.Xie: On integral method, Journal of Xingtai university, Vol.26(2011)No.1, p. 190-192.(in Chinese).

[6] S.S.Mao, Y.M.Cheng, X,L.Pu: Probability Theory and Mathematical Statistics (Second Edition) (Higher Education Press, China, 2010),P. 167-168.(in Chinese).

[7] https://wenku.baidu.com/view/ec5aa9e9d25abe23482fb4daa58da0116d171f5a.html.

[8] L.Wasserman: All of Statistic A Concise Course in Statistical Inference(Springer, Germany, 2003),p.34-34.